REMARKS

Claims 1-21 were pending in this application. A Preliminary Amendment was filed on April 2, 2008 which amended claim 1 and canceled claims 2-21. The Office Action mailed April 7, 2008 rejected claims 1-21. Thus, the Applicant responds herein to the Office Action mailed April 7, 2008 based on the understanding that the Preliminary Amendment was not entered.

In this response, the Applicant has amended claims 1-3, 5-7, 9-10, 12-13, and 15-17, and canceled claims 4, 8, 11, 19, and 20-21. Accordingly, claims 1-3, 5-7, 9-10, and 12-18 remain pending. The Applicant respectfully submits that the present application is in condition for allowance.

Specification

Please amend the title of the application as indicated above. The amendment to the title reflects the current status of the application. The Applicant submits that no new matter has been added.

Claim Rejections

The Office Action rejected claims 1-7, 9-15, 17-18, and 20-21 under 35 U.S.C. 102(e) as being anticipated by U.S. Patent Application Publication Number 2004/0030741 of Wolton et al. (hereinafter "Wolton"). The Office Action also rejected claim 8 under 35 USC 103(a) as being unpatentable over Wolton in view of U.S. Patent Application Publication Number 2004/0064471 of Brown et al. (hereinafter "Brown"), and claims 16 and 19 under 35 USC 103(a) as being unpatentable over Wolton in view of U.S. Patent Application Publication Number 2005/0060295 of Gould et al. (hereinafter "Gould").

Claim 1 has been amended to recite (emphasis added):

1. A computer-implemented method for information retrieval, classification, indexing, and summarization, comprising:
identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors, wherein the identifying includes observing results of a first number of heuristics

run on the collection of hyperlinked documents and related hyperlinks, wherein the first number of heuristics includes identifying at least one of: similar creation dates and similar last-modified dates;

analyzing the content and structure of the compound document to find a preferred entry point for the compound document, wherein the analyzing includes observing results of a second number of heuristics run on the compound document and related hyperlinks, wherein the analyzing includes combining the results of the second number of heuristics run on various hyperlinked documents of the compound document, wherein the results of the second number of heuristics include numerical scores and the combining includes a weighted averaging of the numerical scores into an overall score, and wherein a maximum overall score determines the preferred entry point;

processing the compound document as a whole, including at least one of indexing, classification, and retrieval; and

processing the compound document from the entry point, including at least one of creating at least one of presentation of results from retrieval, summarization, and classification.

The Applicant respectfully submits that claim 1, considered as a whole, is patentable over Wolton, Brown and Gould.

For example, claim 1 requires, *inter alia*, "identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors." Wolton, Brown and Gould do not disclose this limitation.

Wolton describes "A modular intelligent personal agent system...for search, navigation, control, retrieval, analysis, and results reporting on networks and databases." (Wolton, Abstract.) Wolton describes that what distinguishes it "from other search and retrieval agent systems available for application to the World Wide Web, is that it provides a open ended flexible agent creation and configuration tool that does not require any programming experience to use, and thereby permits non-programmer users the ability to generate sophisticated web search and retrieval agents and suites of agents." (Wolton, ¶0169.) Paragraph 152 of Wolton, cited by the Office on page 2 of the Office Action, describes a visual representation of hyperlinked documents as nodes and of the links between hyperlinked documents as connector lines:

> [0152] A client-side or server-side software application retrieves hypertext documents executing a user-selected search algorithm, which search results are displayed in several alternate three-dimensional graphical visualization formats.

> Hypertext documents are displayed as symbol or thumbnail nodes with connector lines representing links between the web documents, and nodes and connector lines are color coded for the user according to the truth of search terms tested for those documents, or according to domain type, link density, or metric counts. Different symbols can represent search and Boolean determination status, document type, and thumbnails can represent reductions of the whole or portions of a document page or type document found.

Paragraph 152 of Wolton does not describe identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors.

Paragraphs 363-367 of Wolton, cited by the Office on page 2 of the Office Action, also do not describe identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors. Rather, paragraphs 363-367 of Wolton describe three "Navigation modes" used by Wolton to determine which links on an HTML page to follow during a search. Wolton states "Each Navigation mode provides a different qualitative type of pruning of the page hyperlink investigation process, thereby providing different capability of the search to follow all links associated with a page or only to follow selective links associated with a page." (Wolton, ¶367.)

Paragraph 832 of Wolton, cited by the Office on page 2 of the Office Action, also does not describe identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors. Rather, paragraph 832 of Wolton describes that Wolton's system provides a second agent with a URL or local HTML page when a first agent triggers the second agent but does not provide the second agent with any URL starting page information.

Brown also does not describe "identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors." Brown describes "A method for presenting content from the page in a distributed database." (Brown, Abstract.) In Brown's a preferred embodiment, "a server receives a request from a client for a page from the database wherein the page has a plurality of links to linked pages in the database. The server retrieves the page

and generates a set of thumbnails of the linked pages in the database. The server then sends the page and the set of thumbnails to the client." (Brown, Abstract.)

Gould also does not describe "identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors." Gould "relates to network communication systems, and more particularly to statistical classification of network data for signature-based security and quality-of-service." (Gould, ¶0002.) Gould describes "A network data classifier [that] statistically classifies received data at wire-speed by examining, in part, the payloads of packets in which such data are disposed and without having a priori knowledge of the classification of the data." (Gould, Abstract.)

Wolton, Brown and Gould do not disclose the limitation of claim 1 of "identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors."

Furthermore, claim 1 requires not only "identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors," but that "the identifying includes observing results of a first number of heuristics run on the collection of hyperlinked documents and related hyperlinks, wherein the first number of heuristics includes identifying at least one of: similar creation dates and similar last-modified dates."

As the Office Action admits, Wolton does not disclose wherein the heuristic includes identifying at least one of: similar creation dates and similar last-modified dates. (Office Action, p. 7.) The Office Action cites Brown as disclosing that a page has a plurality of links to linked pages in a database and that web page information such as creation dates can be searched. (Office Action, p. 7.) The Office Action asserts that it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Wolton with Brown to better identify the searching pages.

First, Brown's description of Brown's use of creation dates does not suggest to one of ordinary skill identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors, wherein the identifying includes observing results of a first number of heuristics

run on the collection of hyperlinked documents and related hyperlinks, wherein the first number of heuristics includes identifying at least one of: similar creation dates and similar last-modified dates. Brown describes:

> Negative preferences may be content related where the user has indicated-key words or subject matter which is not wanted such as adult oriented material. Other examples of negative preferences include or relate to the size of the web page; avi's; music; number of links; number of images; total size of images; JavaScript presence; Java Applet presence; domain name suffix; author; and date of information, i.e. less than seven days old. If such unwanted material or characteristics are present on the web page, then the appearance of the currently viewed web page is altered to reflect such information (step 1225). Examples of such modification include presenting an image of a circle with a line through it next to the link to indicate that the associated web page contains unwanted characteristics. (Brown, ¶0062, emphasis added.)

> If the web page does not contain negative preferences, then the web page is parsed to determine if it contains more than a threshold amount of positive preferences (step 1230). Positive preferences (or criteria) are preferences that the user desires in a web page. The positive preferences may relate to content and key words or it can relate to characteristics about the web page itself such as date of creation, author, etc. Thus, the same kinds of information can be searched for whether desired (positive preferences) or unwanted (negative preferences). Other examples of user specified criteria or preferences include determining the speed of the download for a particular linked page or whether a web page is secure (these could also be included as negative criteria as well). If the amount of positive preferences exceeds a threshold (step 1230), then the appearance of the current web page is modified to indicate such information (step 1235). (Brown, ¶0063, emphasis added.)

Accordingly, Wolton in view of Brown does not disclose or suggest identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors, wherein the identifying includes observing results of a first number of heuristics run on the collection of hyperlinked documents and related hyperlinks, wherein the first number of heuristics includes identifying at least one of: similar creation dates and similar last-modified dates.

Second, even if the Office Action's statement were true that it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Wolton

with Brown to better identify the searching pages, Wolton and Brown still fail to render obvious wherein the identifying includes observing results of a first number of heuristics run on the collection of hyperlinked documents and related hyperlinks, wherein the first number of heuristics includes identifying at least one of: similar creation dates and similar last-modified dates. This is because the identifying claimed is not identifying the searching pages, as the Office Action asserts, but rather identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors.

Gould also does not describe "the identifying includes observing results of a first number of heuristics run on the collection of hyperlinked documents and related hyperlinks, wherein the first number of heuristics includes identifying at least one of: similar creation dates and similar last-modified dates." As discussed above, Gould describes "A network data classifier [that] statistically classifies received data at wire-speed by examining, in part, the payloads of packets in which such data are disposed and without having a priori knowledge of the classification of the data." (Gould, Abstract.)

Furthermore, Wolton, Brown and Gould also do not disclose the limitation of claim 1 of "analyzing the content and structure of the compound document to find a preferred entry point for the compound document, wherein the analyzing includes observing results of a second number of heuristics run on the compound document and related hyperlinks, wherein the analyzing includes combining the results of the second number of heuristics run on various hyperlinked documents of the compound document, wherein the results of the second number of heuristics include numerical scores and the combining includes a weighted averaging of the numerical scores into an overall score, and wherein a maximum overall score determines the preferred entry point."

The Office Action cites Wolton at paragraphs 662-663 and 800 as disclosing analyzing the content and structure of the compound document to find a preferred entry point for the compound document. The Applicant respectfully submits that Wolton does not disclose analyzing the content and structure of the compound document to find a preferred entry point for the compound document. Wolton at paragraphs 662-663

describes how Wolton's system works when nodes that represent HTML documents in Wolton's visual display are clicked:

> [0662] For example, clicking once on a highlighted node shows in the pop-up window a web page thumbnail picture or the first captured textual data set corresponding to the page node discovered Boolean Match. Clicking again on the same node without changing nodes, will show the text label name for the first file name under image or custom capture extension set up for the agent. If the file extension refers to a web page active file, such as JPEG, GIF, or WAV, or MOV file for example, the file will display or play in the pop up window.
>
> [0663] If the web document custom captured is not a standard browser displayed file content, clicking will display the name of the file, which then clicking again will launch the local application that displays or plays the file. The click selection order of pop-up display may be configured to show:

Wolton at paragraph 800 describes starting an agent or starting an application program using a batch file:

> [0800] Both the On TRUE and On FALSE text entry lines can individually specify more than one agent to start, or can specify the computer to launch any application program, or combination thereof of agents, application programs, using a batch file. The "Browse" buttons associated with each On TRUE and On FALSE sub-window areas can be used to capture a file name and path on the users computer that are to be executed.

Wolton does not disclose or describe analyzing the content and structure of the compound document to find a preferred entry point for the compound document. Moreover, as discussed above, Wolton does not disclose the compound document stated in claim 1, and accordingly would not disclose analyzing the content and structure of the compound document to find a preferred entry point for the compound document.

Furthermore, as the Office Action admits, Wolton does not disclose numerical scores and the combining includes a weighted averaging of the numerical scores into an overall score, and the maximum overall score determines the preferred entry point. (Office Action, p. 8.) The Office Action cites Gould as disclosing at paragraphs 54, 60 and 65 overall score and at paragraphs 56, 59-60 and 64 weight. (Office Action, p. 8.)

The Office Action asserts that it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Wolton with Gould to better analyze the data.

The Applicant respectfully asserts that, irregardless of whether Wolton with Gould would result in better analysis of data, Wolton in view of Gould still does not render claim 1 obvious. Claim 1 requires, *inter alia*, that the analyzing includes combining the results of the second number of heuristics run on various hyperlinked documents of the compound document, wherein the results of the second number of heuristics include numerical scores and the combining includes a weighted averaging of the numerical scores into an overall score, and wherein a maximum overall score determines the preferred entry point. Gould at paragraphs 54, 60 and 65 describe functions and vectors used to identify or determine to which class data or a data packet belongs. Gould at 56, 59-60 and 64 describe weight vectors. Gould does not describe combining the results of the second number of heuristics run on various hyperlinked documents of the compound document wherein the combining includes a weighted averaging of the numerical scores into an overall score, and wherein a maximum overall score determines the preferred entry point for the compound document.

Brown also does not describe the limitations of missing from Wolton and Gould. Brown describes that "a server receives a request from a client for a page from the database wherein the page has a plurality of links to linked pages in the database. The server retrieves the page and generates a set of thumbnails of the linked pages in the database. The server then sends the page and the set of thumbnails to the client." (Brown, ¶0010, emphasis added.) Brown describes "a method of browsing the Internet. A server receives user criteria and a request for a page from the Internet from a client. The server retrieves the page and parses the page for a set of links to a set of linked web pages. The server then retrieves the set of linked pages and parses the set of linked pages for user selected criteria. Responsive to finding the user criteria on a linked page within the set of linked pages, the server modifies the page to indicate the presence of the user criteria on the linked page and sends a modified page to the client." (Brown, ¶0011, emphasis added.)

Brown does not describe analyzing the content and structure of the compound document to find a preferred entry point for the compound document, wherein the

analyzing includes observing results of a second number of heuristics run on the compound document and related hyperlinks, wherein the analyzing includes combining the results of the second number of heuristics run on various hyperlinked documents of the compound document, wherein the results of the second number of heuristics include numerical scores and the combining includes a weighted averaging of the numerical scores into an overall score, and wherein a maximum overall score determines the preferred entry point.

Accordingly, the Applicant respectfully submits that claim 1, considered as a whole, is patentable over Wolton, Brown, and Gould.

Claims 2-3, 5-7, 9-10, 12-13, and 15-17 each depend directly or indirectly from claim 1. Therefore, claims 2-3, 5-7, 9-10, 12-13, and 15-17 are patentable over Wolton, Brown and Gould for at least similar reasons. Claims 4, 8, 11, 19, and 20-21 are canceled in this response.

Accordingly, the Applicant respectfully requests withdrawal of the rejections of claims 1-7, 9-15, 17-18, and 20-21 under 35 U.S.C. 102(e), withdrawal of the rejections of claims 8, 16 and 19 under 35 USC 103(a).

CONCLUSION

The Applicant respectfully submits that the present application is in condition for allowance.

In the event a telephone conversation would expedite the prosecution of this application, the Examiner is invited to call the undersigned at (408) 927-3372. Although no fee is believed to be due, the Commissioner is authorized to charge any such fees in connection with the filing of this paper to Deposit Account No. 09-0441 (Order No. ARC920030028US1).

Respectfully submitted,

Date: July 8, 2008

By: /Van N. Nguy/
Van N. Nguy
Reg. No. 55,851
Intellectual Property Law
IBM Almaden Research Center
650 Harry Road
San Jose, California 95120
Telephone: (408) 927-3372
Facsimile: (408) 927-3375